UNIVERSITY OF
ILLINOIS LIBRARY
AT URBANA-CHAMPAIGN
STACKS

Digitized by the Internet Archive
in 2011 with funding from
University of Illinois Urbana-Champaign

http://www.archive.org/details/marketmodelpoten606coll

330
B385
no 606
Cop. 2

082-88

Faculty Working Papers

THE MARKET MODEL: POTENTIAL FOR ERROR

Dennis J. Collins, Assistant Professor
Department of Accountancy
James C. McKeown, Professor, Department
of Accountancy

#606

College of Commerce and Business Administration
University of Illinois at Urbana-Champaign

FACULTY WORKING PAPERS

College of Commerce and Business Administration

University of Illinois at Urbana-Champaign

September 17, 1979

Not to be quoted without the authors' permission.

THE MARKET MODEL: POTENTIAL FOR ERROR

Dennis J. Collins, Assistant Professor
Department of Accountancy
James C. McKeown, Professor, Department
of Accountancy

#606

Summary:

The purpose of this study is to examine several alternative forms of the capital asset pricing model for specification errors. Based upon an analytical examination of the impact of specification error upon ordinary least squares estimates of a linear model, an empirical test for specification error was developed. Based upon the results of the test, the continuous time versions of the model are more appropriate than the discrete versions of the model for market based research.

The Capital Asset Pricing Model of Sharpe (1964), Litner (1965) and Mossin (1966) has been the subject of much discussion and research. Attention has centered on the ability of the model to represent the relationship, in equilibrium, between the return of an individual security and the return of the market portfolio (consisting of all assets, each in proportion to total value outstanding).

A major criticism of the Capital Asset Pricing Model involves the simplicity of the equilibrium relationship that it posits. Those who have criticized the specification of the market model suggest that additional factors must be considered. Sharpe (1977) suggests that anticipations about key economic variables should be included. Similarly, Turnbull (1977) relates the systematic risk of the firm to the responsiveness between the firm specific and the economic variables with the market portfolio. Since only limited information exists concerning the nature of the additional factors (Lee and Vinso (1979) and Rosenberg (1974)), the use of models employing the richer specifications is necessarily limited. The potential misspecification of the capital asset pricing model and the resulting bias in parameter estimates has far reaching consequences on security market research.

The purpose of this study is to test several alternative specifications of the capital asset pricing model, for monthly returns, for exclusion of relevant explanatory variables. The set of firms used for the analysis are all the firms on the CRSP (Center for Research in Security Prices) data file having complete monthly return data for the period January, 1970 through December, 1978. Specifically the question examined is: Does the exclusion of potentially relevant explanatory variables from

the capital asset pricing model result in a biased set of security return predictions?

Relatively common specifications of the capital asset pricing model are employed as objectively as possible in order to isolate the possible bias due to the omission of relevant explanatory variables. Undoubtedly, a more powerful discrimination of specification errors could be made assuming specific alternatives. The purpose here, however, is to assess whether model misspecification poses a limitation on the current specification of the capital asset pricing model for security price research.

The Misspecification Problem

Traditional market-based research uses some variant of the Sharpe (1964), Litner (1972), Mossin (1966) capital asset pricing model to assess the information effects of an accounting event. Under this approach, the researcher uses estimates of the parameters of the market model and values of the market return to form predictions of firm security return behavior for a period which is independent of the period used to estimate the model parameters. The predicted returns are used along with the actual returns for the period to form return residuals. As Collins and Dent (1978) note, there are several methods used to aggregate and analyze the return residuals.[1] A fundamental assumption employed in the methods used to analyze return residuals is that the ordinary least squares estimates of the model parameters are unbiased. This assumption may not hold

[1] For a review of the methods used to test return-residuals and the related econometric problems see Collins and Dent (1978).

if the form of the market model employed is misspecified by excluding relevant explanatory variables.

The subject of misspecification in linear statistical models is treated in the traditional econometric literature (Johnston (1972), Theil (1971)). However, the following brief scenario may provide insight into the effect on parameter estimates of excluding explanatory variables. To facilitate the explanation, $\underset{\sim}{y}$ will denote an $n \times 1$ vector of observations of the dependent variable, X_1 will denote an $n \times k_1$ matrix of observations on known explanatory variables, X_2 will denote an $n \times k_2$ matrix of observations on the unknown explanatory variables, $\underset{\sim}{\varepsilon}$ will denote an $n \times 1$ vector of disturbance terms which are normal $(0,\sigma^2)$ i.i.d., $\underset{\sim}{B}_1$ denotes the $k_1 \times 1$ vector of coefficients of X_1 and $\underset{\sim}{B}_2$ denotes the $k_2 \times 1$ vector of coefficients of X_2.

Suppose that the correct specification of the dependent variable, $\underset{\sim}{y}$, is:

$$(1) \qquad \underset{\sim}{y} = X_1\underset{\sim}{B}_1 + X_2\underset{\sim}{B}_2 + \underset{\sim}{\varepsilon}$$

where X_1, X_2, $\underset{\sim}{B}_1$, $\underset{\sim}{B}_2$, $\underset{\sim}{\varepsilon}$ are defined as above. However, the variables composing the matrix X_2 are not known. This may, in fact, be analogous to the situation facing accounting researchers doing market based research if the theoretical arguments presented in Sharpe (1977) or Turnbull (1977) adequately describe the investment environment. Under these conditions, the researcher may decide to estimate the model conditional upon the known explanatory variables. This results in the following specification:

(2) $\underset{\sim}{y} = X_1\underset{\sim}{B}_1 + \underset{\sim}{\varepsilon}^*$

where $\underset{\sim}{\varepsilon}^*$ is a vector of disturbance terms. It readily follows from (1) that $\underset{\sim}{\varepsilon}^*$ must have the following form:

(3) $\underset{\sim}{\varepsilon}^* = X_2\underset{\sim}{B}_2 + \underset{\sim}{\varepsilon}$.

Further, it may be shown that $\underset{\sim}{\varepsilon}^*$ is distributed $N(X_2\underset{\sim}{B}_2, \sigma^2 I)$.

The estimates of the coefficients for equation (2) using ordinary least squares have the following form, ^, denotes estimate:

(4) $\hat{\underset{\sim}{B}}_1 = (X_1' X_1)^{-1} X_1' \underset{\sim}{y}$,

(5) $\hat{\sigma}^2 = (1/n-k_1)\ \hat{\underset{\sim}{\varepsilon}}^{*\prime}\ \hat{\underset{\sim}{\varepsilon}}^*$.

Unlike their counterparts from a correctly specified model, the estimates of $\hat{\underset{\sim}{B}}_1$ and $\hat{\sigma}^2$ are biased. That is, the expected value of the estimated parameter does not equal the population parameter being estimated. The expected value of $\hat{\underset{\sim}{B}}_1$ and $\hat{\sigma}^2$ are given below:

(6) $E(\hat{\underset{\sim}{B}}_1) = \underset{\sim}{B}_1 + (X_1' X_1)^{-1} X_1' X_2 \underset{\sim}{B}_2$,

(7) $E(\hat{\sigma}^2) = \sigma^2 + (1/n-k_1)(\underset{\sim}{B}_2' X_2'\ M\ X_2\underset{\sim}{B}_2)$,

where $M = (I - X_1(X_1' X_1)^{-1} X_1')$. From equation (6) and equation (7) it is clear that the bias of $\hat{\underset{\sim}{B}}_1$ is equal to $(X_1' X_1)^{-1} X_1' X_2 \underset{\sim}{B}_2$ and that the bias in $\hat{\sigma}^2$ is equal to $(1/n-k_1)(\underset{\sim}{B}_2' X_2'\ M\ X_2\underset{\sim}{B}_2)$.

As noted in Johnston (1972, p.168), the bias in the estimated coefficients and the residual variance depends upon the relationship between the excluded variables and the included variables. In order

to examine the bias in the estimates further, it is necessary to specify the relationship between the omitted variables and the included variables. For the sake of generality, two extreme relationships are examined--X_2 is orthogonal to X_1, then $X_1'X_2 = \phi_{k_1,k_2}$ where ϕ_{k_1,k_2} is a $k_1 \times k_2$ matrix composed of zeros. If this is the case, then the estimate of B_1 is unbiased. However the bias in the estimate of $\hat{\sigma}^2$ is at a maximum.[2] If X_2 is a linear combination of X_1, then the situation reverses. The $\hat{\sigma}^2$ will be unbiased and the bias in the estimate of $\hat{B}_1$ will be at a maximum.[3] It is unlikely to expect either of the two extreme positions examined to hold in practice. As a result, if one excludes relevant variables from the capital asset pricing model, the estimates of all the parameters in the model will be biased. In particular the estimate of the residual variance will be biased upward.

---

[2]This follows easily from an examination of the bias component in $\hat{\sigma}^2$. Recall that the bias of $\hat{\sigma}^2$, denoted $b(\hat{\sigma}^2) = (1/n-k_1)(B_2'X_2' M X_2B_2)$ where $M = (I-X_1(X_1' X_1)^{-1} X_1')$. Thus it follows that:

$$b(\hat{\sigma}^2) = (1/n-k_1)(B_2'X_2'X_2B_2 - B_2'X_2'X_1(X_1' X_1)^{-1} X_1X_2B_2) .$$

But $B_2'X_2'X_1(X_1' X_1)^{-1} X_1X_2B_2$ is a quadratic form and is always greater than or equal to zero. Similarly, $B_2'X_2'X_2B_2$ is a quadratic form and also greater than or equal to zero. But in the case under consideration $X_1'X_2 = \phi_{k_1,k_2}$ hence $b(\hat{\sigma}^2) = (1/n-k_1)(B_2'X_2'X_2B_2)$ which is greater than $b(\hat{\sigma}^2)$ where $X_1'X_2 \neq \phi_{k_1,k_2}$.

[3]In this case, X_2 is an exact linear combination of X_1. Thus there exists a $k_1 \times k_2$ matrix C such that $X_1C = X_2$. Assuming that X_1 meets the usual rank conditions, $C = (X_1' X_1)^{-1} X_1'X_2$. Substituting $C = (X_1' X_1)^{-1} X_1'X_2$ into the expressions for the bias of $\hat{B}_1$ and $\hat{\sigma}^2$ yields $b(\hat{B}_1) = CB_2$ and $b(\hat{\sigma}^2) = 0$.

The biases in the estimated parameters, aside from the obvious implications on the evaluation of model performance, also affect the nature of the security return residuals used to assess the impact of accounting information. Suppose that the model specified in equation (2) is used to form predictions of security returns of a firm for some period independent of the period used to estimate the model parameters. Let $\underset{\sim}{y}_H$, X_{1H}, X_{2H}, $\underset{\sim}{\varepsilon}_H$ represent the values of the variables during the prediction period. From equation (1), $\underset{\sim}{y}_H$ is related to X_{1H} and X_{2H} as follows:

$$(8) \qquad \underset{\sim}{y}_H = X_{1H}\underset{\sim}{B}_1 + X_{2H}\underset{\sim}{B}_2 + \underset{\sim}{\varepsilon}_H \ .$$

The prediction of $\underset{\sim}{y}_H$, denoted $\hat{\underset{\sim}{y}}_H$, will only depend upon $\hat{\underset{\sim}{B}}_1$ and X_{1H} and may be expressed as follows:

$$(9) \qquad \hat{\underset{\sim}{y}}_H = X_{1H}\hat{\underset{\sim}{B}}_1 \ .$$

The expression for the return residuals, $\hat{\underset{\sim}{u}}_H$, follows directly from equation (8) and equation (9):

$$(10) \qquad \hat{\underset{\sim}{u}}_H = \underset{\sim}{y}_H - \hat{\underset{\sim}{y}}_H = X_{1H}(\underset{\sim}{B}_1 - (X_1' X_1)^{-1} X_1'\underset{\sim}{y}) + X_{2H}\underset{\sim}{B}_2 + \underset{\sim}{\varepsilon}_H \ .$$

The distributional properties of $\hat{\underset{\sim}{u}}_H$ are easily derived from equation (10). Let $E(\cdot)$ denote the expectation operator, the first and second moments of $\hat{\underset{\sim}{u}}_H$ are:

$$(11) \qquad E(\hat{\underset{\sim}{u}}_H) = (X_{2H} - X_{1H}(X_1' X_1)^{-1} X_1'X_2)\ \underset{\sim}{B}_2 \ ,$$

$$(12) \qquad E(\hat{\underset{\sim}{u}}_H - E(\hat{\underset{\sim}{u}}_H))(\hat{\underset{\sim}{u}}_H - E(\hat{\underset{\sim}{u}}_H))' = \sigma^2[I_H + X_{1H}(X_1' X_1)^{-1}X_{1H}'] \ ,$$

where I_H denotes an $H \times H$ identity matrix.

Equation (11) clearly indicates that the exclusion of relevant variables may result in a predictor which is biased. The bias in the resulting residual security returns is related to the bias in the coefficient estimate, $\hat{B}_1$. As before, the exact nature of the bias depend upon the relationship between X_2 and X_1. If X_2 is orthogonal to X_1, the bias is equal to $X_{2H}B_2$. Where as if X_2 is an exact linear combination of X_1 the bias is equal to zero.

The Current Study

The CRSP data file contained 1001 companies listed on the New York Stock Exchange for which complete, monthly security return data was available for the period 1970 through 1978. The 1001 companies formed an initial sample for testing the alternative specifications of the capital asset pricing model described below. However, the number of companies used in the final specification test varied with the model form specified. For each model specification, only those companies for which the computed regression coefficients were found to be stable over the period 1970-1972 and 1976-1978 were used.[4] The analysis of covariance test (Wilson (1978)) at the 5% level of confidence was used to assess stability.

Six alternative specifications of the capital asset pricing model were used. These were:

(1) $r_t = \alpha + Br_{mt} + \varepsilon_{1t}$

(2) $(r_t - r_{ft}) = \alpha + B(r_{mt} - r_{ft}) + \varepsilon_{2t}$

(3) $(r_t - r_{ft}) = B(r_{mt} - r_{ft}) + \varepsilon_{3t}$

(4) $\ln(1+r_t) = \alpha + B \ln(1+r_{mt}) + \varepsilon_{4t}$

[4] It is recognized that the use of the same test period for each company could bias the results of the significance tests below, where each security is assumed independent. However, if the model is correctly specified (the null hypothesis) this effect should be minimal.

$$(5)\quad \ln(1+(r_t-r_f)) = \alpha + B \ln(1+(r_{mt}-r_{ft})) + \varepsilon_{5t}$$

$$(6)\quad \ln(1+(r_t-r_f)) = B \ln(1+r_{mt}-r_{ft})) + \varepsilon_{6t}$$

where r_t is the security return for time t, r_{mt} is a value weighted index of the return on the market portfolio for time t, r_{ft} is the return to a risk free asset at time t which was measured using the monthly return to U.S. Treasury Bills, α and B are the regression coefficients and ε_{it} (i=1,2,...,6) are the regression disturbance terms.

The period January 1970 through December 1978 was partitioned into two subintervals. One of the subintervals was used to estimate the parameters of each of the five models specified. The second subinterval, which is independent of the estimation interval, was used to form security return residuals. The estimation interval consisted of 72 months (January 1970-December 1972 and January 1976-December 1978). The prediction interval consisted of 36 months (January 1973-December 1975). For the companies that had a stable regression relationship over the estimation period, the vector of return residuals, $\hat{u}_H$, was calculated using the estimated coefficients from the regressions and the data of the prediction period.

Test Results

The sets of return-residuals for each model were used to test two null hypotheses concerning model specification errors. One hypothesis involved the sums of the return residuals and was designed to detect possible bias in the return residuals due to the exclusion of relevant explanatory variables. A second hypothesis concerned a test of the independence of the error term and the exogenous variable for each model specified.

Under the first hypothesis, if the model is correctly specified, then the return residuals have expectation zero. This also implies that the sum of the return residuals for the 36 month prediction period have an expectation of zero. This amount for each company m and asset pricing model k (k=1,2,3,4, or 5) is:

$$s_{m,k} = \sum_{i=1}^{36} \hat{u}_{i,m,k}$$

where $\hat{u}_{i,m,k}$ is the return residual of the ith prediction period for firm m using model k. It is assumed that the sign of $s_{m,k}$ is independent across firms. Because specification errors may bias the assessment of the variance of $s_{m,k}$, the sign test is used to test for the location of the median of the distribution of $s_{m,k}$ (Snedecor and Cochran, 1976, p. 211). The null hypothesis tested was that the median of the distribution of the sums of return residuals was zero. The alternative in each case was that the median was not equal to zero. Table 1 presents a summary of the tests.

[Insert Table 1 here]

The null hypothesis was rejected at the 0.05 level for each of the first two models and accepted for the remaining three models. The first model is the familiar market (or diagonal) model. The evidence suggests that the security-return residuals from this model do not have an expectation of zero. Based on the foregoing analysis, biased return residuals are an artifact of biased model parameter estimates. The second model is a variant on the two factor model. The results for this model are similar to the results for the market model--the security return residuals are biased. The three models exhibiting unbiased residuals are all logarithmic forms. The logarithmic form of the model corresponds to monthly rates of return with continuous compounding.

TABLE 1

Specification Test Results

Model	Number of Stable Firms N	The Number of Positive Sums N_p	The Test Statistic* $(\lvert N_p-(1/2)N\rvert-1/2)/\ (.5)^2N$
(1) $r_t = \alpha + Br_{mt} + \varepsilon_{1t}$	890	498	3.5196
(2) $(r_t-r_{ft}) = \alpha + B(r_m-r_{ft}) + \varepsilon_{2t}$	883	489	3.1634
(3) $(r_t-r_{ft}) = B(r_{mt}-r_{ft}) + \varepsilon_{3t}$	903	629	11.7804
(4) $\ln(1+r_t) = \alpha + B\ \ln(r_{mt}) + \varepsilon_{4t}$	870	441	0.3729
(5) $\ln(1+(r_t-r_{ft})) = \alpha + B\ \ln(1+(r_{mt}-r_{ft})) + \varepsilon_{5t}$	866	439	0.3738
(6) $\ln(1+(r_t-r_{ft})) = B\ \ln(1+(r_{mt}-r_{ft})) + \varepsilon_{6t}$	906	477	1.5615

*The test statistic follows the normal distribution. The null hypothesis is rejected at the 0.05 significance level for a test statistic greater than or equal to 1.96.

The second hypothesis tested involved the independence of the disturbance terms and the explanatory variables. For these tests, the statistic of interest for a particular firm m and model k (k=1,2,3,4,6) is:

$$\hat{p}_{m,k} = \frac{1}{36} \sum_{i=1}^{36} (\hat{u}_{i,m,k} - \bar{\hat{u}}_{m,k})(X_{i,m,k} - \bar{X}_{m,k})$$

where $\hat{u}_{i,m,k}$ is the return residual of the ith prediction period for firm m using model k, $X_{i,m,k}$ is the value of the exogenous variable in the ith prediction period for firm m and model k, $\bar{\hat{u}}_{m,k}$ and $\bar{X}_{m,k}$ are the arithmetic means of $\hat{u}_{i,m,k}$ and $X_{i,m,k}$ over the 36 month prediction interval. Fisher's z transformation was applied to $\hat{p}_{m,k}$ to obtain an almost normally distributed variable (Snedecor and Cochran, 1976, p. 185). A test of proportions is used to test the location of the mean of the transformed sample statistics of each model (Snedecor and Cochran, 1976, p. 211). The null hypothesis tested, for each model, was that the proportion of firms falling outside of a 0.0455 level of confidence centered at zero equalled 0.0455. The alternative in each case was that the proportion was not equal to 0.0455. Table 2 presents a summary of the tests.

[Insert Table 2 here]

The null hypothesis was rejected in all cases. This suggests that the explanatory variables and the disturbance term are not independent. A possible explanation for this result is that the coefficients of the explanatory variable are random. The random coefficient version of the single index market model (model 1) has been examined by Fabbozzi and Francis (1978). Their research indicates that the random coefficient

TABLE 2

Tests of Independence

Model	Number of Stable Firms N	Number of Correlations Outside of a 0.0455 Confidence Interval N_0	The Test Statistic* p=0.0455 q=0.9545 $(\lvert N_0-pN\rvert - 1/2)/\sqrt{Npq}$
(1) $r_t = \alpha + Br_{mt} + \varepsilon_{1t}$	890	189	23.8061
(2) $(r_t-r_{ft}) = \alpha + B(r_{mt}-r_{ft}) + \varepsilon_{2t}$	883	182	22.8213
(3) $(r_t-r_{ft}) = B(r_{mt}-r_{ft}) + \varepsilon_{3t}$	903	200	25.2962
(4) $\ln(1+r_t) = \alpha + B \ln(r_{mt}) + \varepsilon_{4t}$	870	182	23.0874
(5) $\ln(1+(r_t-r_{ft})) = \alpha + B \ln(1+r_{mt}-r_{ft})) + \varepsilon_{5t}$	866	176	22.0731
(6) $\ln(1+(r_t-r_{ft})) = B \ln(1+(r_{mt}-r_{ft})) + \varepsilon_{6t}$	906	199	25.2962

*The test statistic follows the normal distribution. The null hypothesis is rejected at the 0.05 significant level for a test statistic greater than or equal to 1.96.

version of the model appears to be the most appropriate specification of the return generating process for the 1965-1971 return generating period. The primary effect of the random coefficients model is that the disturbance term is heteroscedastic. This results in a loss in efficiency of the ordinary least squares estimates. However the ordinary least squares estimates are still unbiased.

Conclusions

In this study, an attempt has been made to establish an empirical basis to help resolve the question of which specification of the capital asset pricing model is appropriate for accounting market based research. The study covers only a nine year period and is limited to the firms having complete return information for the period. The approach has of necessity been one of using rather basic methods to assess, in a broad sense, the predictive qualities of the models.

Even on the broad basis of examination, however, predictions based on the logarithmic forms of the capital asset pricing model appear to be unbiased. Whereas, predictions based upon the discrete models of market behavior appear to be biased. The tests applied, while admittedly weak, indicate that the logarithmic forms of the capital asset pricing model are the more appropriate forms for use in accounting research.

References

Collins, D. W. and W. T. Dent, "Econometric Testing Procedures In Market Based Accounting Research," Mimeo, Department of Accounting, University of Iowa, 1978.

Fabozzi, F. J. and J. C. Francis, "Beta as a Random Coefficient," Journal of Financial and Quantitative Analysis, March 1978.

Johnston, J., Econometric Methods (New York: McGraw-Hill Book Company, 1972).

Lee, C. F. and J. D. Vinso, "The Single vs. Simultaneous Equation Model In Capital Asset Pricing: The Role of Firm Related Variables," Journal of Business Research, forthcoming.

Litner, J., "The Valuation of Risk Assets and Capital Market Equilibrium," in Studies in the Theory of Capital Markets (M. C. Jensen, ed.), New York: Praeger, 1972.

Mossin, J., "Equilibrium in a Capital Asset Market," Econometrica, October, 1966.

Rosenberg, B., "Extra-Market Components of Covariance in Security Returns," Journal of Financial and Quantitative Analysis, March 1974.

Sharpe, W. F., "Capital Asset Prices: A Theory of Market Equilibrium Under Conditions of Risk," Journal of Finance, September 1964.

Sharpe, W. F., "The Capital Asset Pricing Model A "Multi" Beta Interpretation," in Financial Decision Making Under Uncertainty (H. Levy and M. Sarnat, eds.), New York: Academic Press, 1977.

Snedecor, G. W. and W. G. Cochran, Statistical Methods (Ames: The Iowa State University Press, 1976).

Theil, Henri, Principles of Econometrics (New York: John Wiley and Sons, Inc., 1971).

Turnbull, S. M., "Market Value and Systematic Risk," The Journal of Finance, Vol. XXXII, No. 4 (September 1977).

Wilson, A. L., "When is the Chow Test UMP," The American Statistician, Vol. 32 (May 1978).

Faculty Working Papers

AN ANALYSIS OF FACTORS THAT INFLUENCE AGGREGATE STOCK MARKET VOLATILITY

Frank K. Reilly, Professor, Department of Finance
John M. Wachowicz, Jr., University of Tennessee

#607

College of Commerce and Business Administration
University of Illinois at Urbana-Champaign

BOUND BY
HW

UNIVERSITY OF ILLINOIS-URBANA
3 0112 060296339